Expert Systems With Applications 178 (2021) 115013

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Robust automated Parkinson disease detection based on voice signals with transfer learning

Onur Karaman [a], Hakan Çakın [b], Adi Alhudhaif [c], Kemal Polat [d,*]

[a] Akdeniz University, Vocational School of Health Services, Department of Medical Imaging Techniques, Antalya 07070, Turkey
[b] Akdeniz University, Faculty of Medicine, Department of Neurosurgery, Antalya 07070, Turkey
[c] Department of Computer Science, College of Computer Engineering and Sciences in Al-kharj, Prince Sattam bin Abdulaziz University, P.O. Box 151, Al-Kharj 11942, Saudi Arabia
[d] Department of Electrical and Electronics Engineering, Bolu Abant Izzet Baysal University, Bolu, Turkey

ARTICLE INFO

Keywords:
Parkinson's disease (PD)
Acoustic sensing
Convolutional neural network (CNN)
Transfer Learning
Voice signal

ABSTRACT

Parkinson's disease (PD) is a progressive-neurodegenerative disorder that affects more than 6 million people around the world. However, conventional techniques for PD detection are often hand-crafted, in which special expertise is needed. In this study, considering the importance of rapid diagnosis of the disease, it was aimed to develop deep convolutional neural networks (CNN) for automated PD identification based on biomarkers-derived voice signals. The developed CNN methods consisted of two main stages, including data pre-processing and fine-tunning-based transfer learning steps. To train and evaluate the performance of the developed model, datasets were collected from the mPower Voice database. SqueezeNet1_1, ResNet101, and DenseNet161 architectures were retrained and evaluated to determine which architecture can classify frequency-time information most accurately. The performance results revealed that the proposed model could successfully identify the PD with an accuracy of 89.75%, sensitivity of 91.50%, and precision of 88.40% for DenseNet-161 architecture identified as the most suitable fine-tuning architecture. The results revealed that the proposed model based on transfer learning with a fine-tuning approach provides an acceptable detection of PD with an accuracy of 89.75%. The outcomes of the study confirmed that by integrating the developed model into smart electronic devices, it will be able to develop alternative pre-diagnosis methods and will assist the physicians for PD detection during the in-clinic assessment. The success of the proposed model would imply an enhancement in the life quality of patients and a cost reduction for the national health system.

1. Introduction

Parkinson's disease, initially called shaking palsy, is a neurological disorder caused by the breakdown of cells in the area of the midbrain called substantia nigra- "the movement control center" of the brain (Fig. 1) that are responsible for producing dopamine (Almeida et al., 2019; Parkinson, 2002; Pereira et al., 2019). Loss of dopamine production causes neurons to fire out-of-control movements called hypokinetic movement disorder (Pahuja and Nagabhushan, 2018). The disease manifests itself with slow movements, tremors, differences in gait, muscle rigidity, besides the non-motor group symptoms such as dementia, depression, anxiety, the disorder of sleep, and slow thinking (Almeida et al., 2019; Berus, Klancnik, Brezocnik, & Ficko, 2019; Skodda, 2011; Bugalho and Viana-Baptista, 2013). Although it is mostly seen in people after 60 years of age, it can also be encountered in the 40 s due to genetic reasons. Living with this disease threatens people's ordinary lives severely (Berus et al., 2019; Reeve, Simcox, & Turnbull, 2014). Thus, the diagnosis of PD is of critical importance for improving the quality of daily activities and extending the lifespan of the patients (Erdogdu Sakar, Serbes, & Sakar, 2017). Since the symptoms and progression of the disease significantly may vary for each person, it is quite difficult to determine when PD symptoms evolve and how they will impact the patient's life over time time (Skodda, 2011; Bugalho and Viana-Baptista, 2013).

The diagnosis of Parkinson's is generally made by the physician according to complaints of the patient and the neurological examination to

* Corresponding author at: Bolu Abant Izzet Baysal University, Faculty of Engineering, Department of Electrical and Electronics Engineering, Bolu, Turkey.
E-mail addresses: aonurkaraman@akdeniz.edu.tr (O. Karaman), hcakin@akdeniz.edu.tr (H. Çakın), A.alhudhaif@psau.edu.sa (A. Alhudhaif), kpolat@ibu.edu.tr (K. Polat).

https://doi.org/10.1016/j.eswa.2021.115013
Received 7 March 2021; Received in revised form 30 March 2021; Accepted 7 April 2021
Available online 13 April 2021

be performed after the disease history (Berus et al., 2019). Although some costly methods like radiological imaging techniques such as computer tomography (CT), X-ray imaging, single-photon emission computerized tomography/dopamine transporter scan, etc., these techniques can effectively detect the PD only if it is spread over the brain (Arena & Stoessl, 2016; Berus, Klancnik, Brezocnik, & Ficko, 2019; Naseer et al., 2020). On the other hand, the finger-touch and handwriting laboratory tests are the novel computer-aided tests applied based on the first symptoms of suspected PD cases (Almeida et al., 2019; Bernardo et al., 2019; de Souza, Alves, Rebouças, Almeida, & Rebouças Filho, 2018; Pereira et al., 2016; Lauraitis, Maskeliūnas, & Damaševičius, 2018; Júnior et al., 2020). In recent years, developing voice tests is a promising research area that processes phonation and voice signals to classify PD-cases. Mechanically, voice production involves lungs air volume and flow, complex structure within the glottis, and its control by laryngeal muscle activation. The vocal cord vibration by laryngeal muscular control provides the fundamental frequency of the voice, vocal intensity, and voice quality (Parkinson et al., 2012). The vocal cords are located in the larynx and give a constricted shape to the airway, and lie in the larynx along the anterior-posterior direction, attaching to the thyroid cartilage from anterior and to the arytenoid cartilages from posterior (Fig. 2). There are three major sound production mechanisms, one of these is a monopole sound source due to the volume of air displaced by vocal cord vibration, the second mechanism is dipole sound source due to the fluctuating force applied by the vocal cords to the airflow, and the third one is a quadrupole sound source due to turbulence developed immediately downstream of the glottal exit (Zhang, 2016). All these mechanisms are mechanical parts of voice production, but there is another part of voice production named the neurological pathway. The neural systems involved in the sensory control of the voice provide the role of auditory feedback in voice control. In particular, the audio-vocal integration and the ability to control fundamental voice frequency (F0) are essential for successful and efficient speech production. The associations of left and right hemisphere cortical nodes with the thalamus that were significantly greater in healthy people are shown of the orientation of speech and vocalization. It is believed the reduced connectivity between the thalamus and superior temporal gyrus cortical nodes is responsible for motor and sensorimotor dysfunction in Parkinson's disease (Van Stan et al., 2017).

It is reported that over 90% of patients with PD have characteristic patterns of atrophy and language disability, which is one of the initial hallmarks of early-stage PD (Sapir, 2014; New et al., 2015; Pawlukowska et al., 2015). The voice turns soft, and the speech turns monotonous and faster. The voice becomes less audible in time; the patient can just whisper in later times of disease. The mumbling can be an early sign of disease. Caekebeke *et al.* reported that the cause of speech disorder is respiratory, articulatory, and phonatory problems (Caekebeke, Jennekens-Schinkel, Van der Linden, Buruma, & Roos,1991). It is estimated that over 90% of patients with Parkinson's disease develop a speech disorder known as hypokinetic dysarthria. The main vocal symptoms of PD are reduced prosody and loudness. Because of the reduced superior temporal gyrus connectivity, these errors not easily correct. There are some studies that provide evidence that Parkinson's disease population exhibits major reductions in left thalamocortical, bilateral putamen, and cortical audio-vocal connectivity in the voice network. To detect these abnormal voice characteristics, some verbal experiments, including running speech experiments and sustained phonation experiments, can be conducted (Dejonckere et al., 2001). These characteristic features in vocal cords such as silent voice, hoarseness, soft and monotonous speech, shortage of air, and tremor of the voice, as described in PD-related hypokinetic dysarthria can be visualized through direct laryngoscopy (Blumin, Pcolinsky, & Atkins, 2004; Hin et al., 2008; Martens, Van Nuffelen, Wouters, & De Bodt, 2016). In voice analysis people suffer from PD perform shorter maximum phonation time, higher voice frequencies, decreased pitch range, and increased phonation threshold pressure (Berus et al., 2019; Chenausky, MacAuslan, & Goldhor, 2011). Although the previous studies suggest the identifiable speech signs onset about 84 months in Parkinson's disease, the recent prospective studies proved that speech signs may detect with objective, reliable measures (Ramig, Halpern, Spielman, Fox, & Freeman,2018).

In most recent years, there has been a steady increase in the number of articles published in the literature using artificial intelligence-based models to diagnose PD and some similar neurological diseases (de Souza et al., 2021; Gupta, Julka, et al., 2018; Gupta, Sundaram, Khanna, Hassanien, & De Albuquerque, 2018). These include speech assessment (Tsanas, Little, McSharry, Spielman, & Ramig, 2012; Millian-Morell et al., 2018; Caliskan, Badem, Basturk, & Yuksel, 2017) gait monitoring (Delrobaei et al., 2018; Xia, Yao, Lu, Zhang, & Cheng, 2018; Xu, He, Zhang, Wang, & Duan, 2018) or tremor assessment. In these studies, various types of input have been used such as hand-written, voice, speech patterns to classify PD hallmarks besides wearable sensors that detect muscular movements (Pereira et al., 2018). While all of these studies are extremely valuable, a majority of them are based on motion sensing and imaging (Choi, Ha, Im, Paek, & Lee, 2017; Singh and Samavedham, 2015) (Table 1). Using non-invasive, easily-obtained, low-cost, patient-generated voice data provides a significant improvement to disease detection. Therefore, recent works have focused to demonstrate the effectiveness of employing machine learning algorithms to classify PD from non-PD using extracted voice features, and

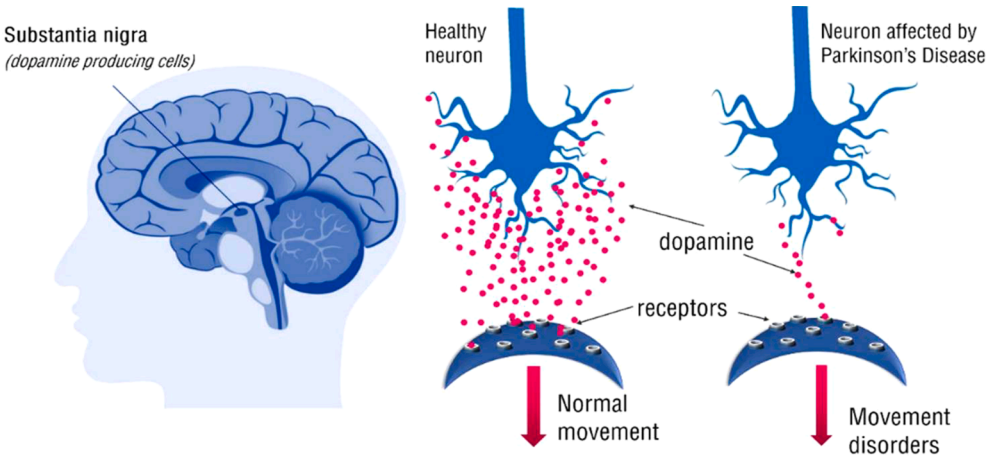

Fig. 1. Illustrative diagram showing a region of the brain responsible for dopamine production and the difference in dopamine levels at a synapse between neurons in a healthy person (left) and a patient with Parkinson's disease (right).

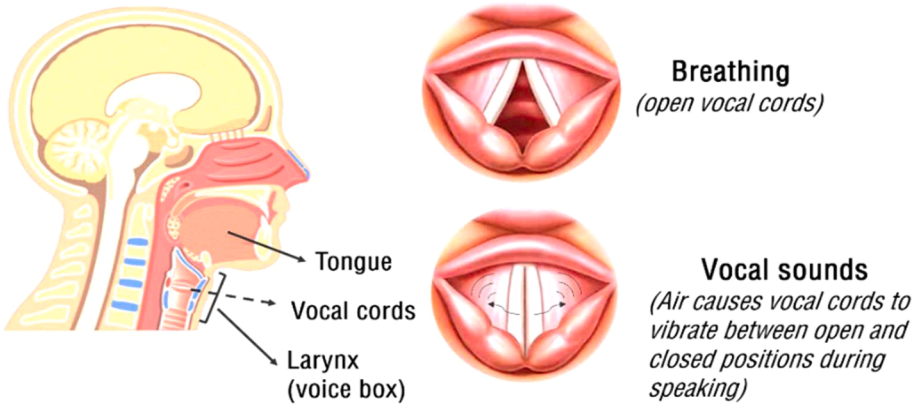

Fig. 2. Illustration of larynx and vocal cords involved in voice production.

Table 1
Summary of deep learning-based models for identifying PD.

Reference	Feature	Method	Data description	Accuracy %
(Pereira et al., 2016)	Pen-based features	CNN	Handwriting dataset	80.2
(Wroge et al., 2018)	Voice features	DNN	Voice dataset	86.0
(Zhang, 2017)	Voice features	DNN	Speech dataset	90.5
(Afonso et al., 2019)	Smartpen based features	CNN	Handwriting data	90.0
(Zhang et al., 2015)	Voice features	DNN	Non-speech body sound	83.3
(Kostikis et al., 2015)	Smartphone-based features	BagDT	Hand tremor data	82.0
(Moetesum et al., 2019)	Automated visual features	CNN	Handwriting data	93.0

DNN; Deep neural network.
BagDT; Bag of decision trees.

most of them reported high accuracy values up to 98%–99% (Wroge et al., 2018; Tsanas et al., 2012; Shirvan and Taham, 2011; Haq et al., 2019; Grósz, Busa-Fekete, Gosztolya, & Tóth, 2015; Vasquez-Correa et al., 2018; Lahmiri & Shmuel, 2019; Upadhya, Cheeran, & Nirmal, 2018; Correia et al., 2019; Mallela et al., 2020). However, the crucial problem is the identity confusion caused by the multiple voice samples collected from the same person. Most of the reported models consider the same voice samples both in the training and the testing data set. As a result, this bias leads to over-optimistic performance results since the model has learned to detect the characteristics of certain speakers and predicted the label in the test set. Therefore, in this work, by considering this problem the voice samples used in training and testing were carefully selected to belong to different people in order to prevent bias and find more realistic approaches.

The direct application of CNN could not be reliable due to the availability of small data size. To overcome this limiting data challenge, transfer learning is promising to tune the already grained storing knowledge on a similar problem (Naseer et al., 2020; Polat et al., 2021). Similar studies based on the detection of PD from voice signals have obtained similar performances by combining feature selection and discrimination algorithms (Sakar and Kursun, 2010; Peker, Sen, & Delen, 2015; Peker, 2016). Moreover, some research declares that a simple transfer learning approach with few parameters can also be competitive, sometimes even better than methods based on deep layers (Naseer et al., 2020; Coates, Ng, & Lee, 2011).

With this motivation, in this study, the authors are dedicated to developing a deep convolutional neural network classifier with a transfer learning-based model for the detection of patients suffering from Parkinson's disease by utilizing the sustained vowels as voice biomarkers. In this study, it is aimed to prevent possible bias that may arise from the dataset and to classify PD with high accuracy by using a more realistic approach. With this scope, firstly Mel-spectrogram was calculated for pre-processing of datasets. The pre-processing approach results in increasing the accuracy of the deep CNN model. Then, in the transfer learning stage, three pre-trained architectures including ResNet101, DenseNet161, and SqueezeNet1_1 were retrained to distinguish PD-case. The proposed CNN model can be used to differentiate participants with PD who exhibit little to no symptoms from healthy controls. This work highlights the potential of voice to be used for the detection of PD and indicates that voice may serve as a deep phenotype for PD, enabling precision medicine by improving the speed, accuracy, accessibility, and cost of PD management. Furthermore, the importance of the proposed model is mainly being an assistant to the neurologists by giving preliminary ideas, additionally helping them to put in the priority order. The obtained digital biomarkers not only help the accurate detection of PD patients but also provide enhanced care by assisting with treatment management. The developed model will not be a primary diagnosis method, by integrating it into smart electronic devices, it will be able to develop alternative pre-diagnosis methods and will assist the physicians.

The novelty of this study is that it offers deeper insight into not only how different feature selection techniques and CNN architectures affect the performance of classification of PD-case but also how the visual patterns of voice images can be utilized in the identification of PD-case.

2. Methods

2.1. Dataset description and data selection

The PD dataset used in this study was collected by the allowance of mPower governance, from the mPower public data portal in which participants self-guide through a visually engaging and robust informed consent process (Sage Bionetworks, 2020). The mPower app available on the Apple App Store (mpower App store, 2020) was launched in March 2015 to evaluate both the severity of PD symptoms and the sensitivity towards the medication in PD. The mPower consists of big data released from seven different modules including demographics, Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) questionnaire, Parkinson's Disease Questionnaire (PDQ), memory, tapping, voice, and walking. Additionally, a mapping of questions asked and variables are available in the demographic survey task database, in which participants responded to the questions about general demographic topics and health history. There are 65,022 voice task data available. In this app, the participants are responsible for stating the time when relative to taking medication, they complete the tasks (e.g., immediately before taking their medication, after taking medication or any time in the day). On the other hand, the participants

who identified themselves as not having a professional diagnosis of Parkinson's disease have been labeled as "the controls" (Fig. 3). Considering the information collected from each task, the obtained data were classified into three groups based on medication time and clinical history as shown in Fig. 3. Firstly, the patients with and without professional PD diagnosis were separated from each other on the mPower database. While data of patients with unknown PD diagnoses were not used, the negative diagnosis data were set as the control group. Furthermore, all data with positive PD diagnoses that had deep brain stimulation surgery or no surgery information were eliminated. Patients who did not have surgery were marked as positive if they did not take medication before recording a voice or did not use any PD medication at all. Since the parkinsonian sound markers in the voices of patients with deep brain stimulation surgery will weaken or disappear completely, this uncertainty was not included in the training data to prevent this uncertainty from decreasing the performance of the model. Moreover, considering the bias problem caused by using identity confounding, during the data classification process, the ID information of the patients take into account and only one voice sample from the same indiviual was selected manually. In addition, only 10-second audio recordings were taken into account to ensure homogeneous feature extraction. After all these elections, there were 23,288 voice records of the control group and 10,589 voice records of the PD positive group. Of the 33,877 sound recordings, 18,660 control group data and 8442 PD positive group data were used for training, while 4628 control group data and 2147 PD positive group data were used for validation. Moreover, totally 400 independent data consisting of 200 PD positive, 200 control group data were used for testing.

The voice activity task is based on a sustained voicing of "aaaah" at a steady loudness level and comfortable pitch for up to 10 s. The obtained data from this task include audio files for both 10 s of phonation and 5 s of a countdown leading up to the activity. All the data in the data set was recorded as lossless using Apple Lossless Audio Codec (ALAC) (ALAC, 2020) with 44.1 kbps bitrate.

The downsampling and reshaping of the data were conducted since the source file contains too large samples. Even if the frequency of voice signal and the tremor are 300–3400 Hz and 5–7 Hz, respectively, the high-frequency bands may play a crucial role in PD identification (Zhang, Wang, Li, & Xu, 2018). Thus, the frequency window was determined as 0 Hz to 22.05 kHz to avoid the data loss. The generalized algortihm for PD forecasting were depicted in Fig. 4.

2.2. Feature extraction

Studies in the literature show that 2-dimensional (2D) visual representations of sound recordings can be classified with high accuracy by CNN networks. Besides, in PD diagnostics, 2D representation of sounds, such as short-time Fourier transform, is successful. Bearing the information in the literature and the importance of computational efficiency in mind, in this study, the discrete cosine transformation (DCT) method was used for feature extraction. After DCT transformation was performed by multiplying the frame buffer selected by the length of 2048 with the Hann window (Hanning, 2020), the frame buffer elements were shifted over the waveform to 512 samples and the process was repeated for the total sample length. It is aimed to prevent spectrum leakage and increase the spectral resolution by using the Hann window.

It is possible to express a time-dependent signal with DCT as the sum of the cosine functions oscillating at different frequencies. The main difference of DCT transformation from Discrete Fourier Transform (DFT) is that it contains only real components. As with DFT, it does not contain imaginary components. For the feature extraction process, type-II DCT transform, generally known as DCT only, is shown with Eq. 1.

$$X_k = \sum_{n=0}^{N-1} x_n cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right] k = 0, \cdots, N-1 \quad (1)$$

After the transformation filterbank matrix created to combine DCT bins into Mel-frequency bins. 320 mels are used for this process. With all of these informations, we have achieved a two-dimensional matrix which holds both time and frequency information about the voice signal. These 2D matrix elements are logarithmically colored with the colormap named "jet" of the matplotlib library. Then, each spectrogram was saved as lossless PNG files in 448px × 448px resolution and this process was

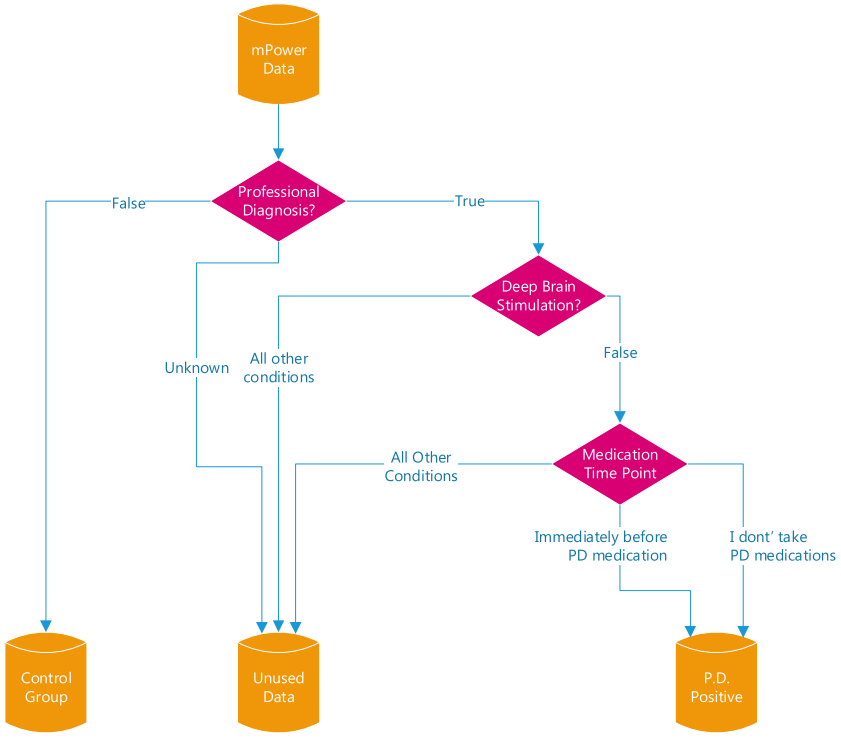

Fig. 3. Classification of data based on medication time and clinical history.

Fig. 4. The generalized algortihm for PD forecasting.

repeated for the entire dataset to train the CNN networks. In Fig. 5, the time domain recording waveforms and time–frequency domain Mel-spectrograms of the control group sample and P.D. positive positive samples are depicted.

2.3. Transfer learning and CNN architectures

Since there is an insufficient amount of training data, to train a CNN network with random weights and biases, pre-trained CNN architectures were retrained and fine-tuned. This approach is known as the transfer learning technique in the literature (Naseer et al., 2020). The transfer learning and optimization of the model were conducted on a personal computer (Ryzen 5 1600 CPU, 16 GB DDR4 RAM, Nvidia RTX 2080 Ti Graphics Card). Fastai (Fastai, 2020) and LibROSA (Librosa, 2020) libraries were used to develop the model. To determine the architecture that can classify time–frequency data in the most accurate way, different architectures including SequeezeNet (Iandola et al., 2016), Residual Neural Network (ResNet) (He, Zhang, Ren, & Sun, 2016), and Dense (DenseNet) (Huang, Liu, Van Der Maaten, & Weinberger, 2017) architectures were trained. These architectures have already been trained by the ImageNet (ImageNet, 2020) database and are able to make a classification in 1000 different classes. Hence, they were retrained to distinguish PD-positive and PD-negative cases by transfer learning method. Whereas ResNet and DenseNet architectures are dense architectures, SqueezeNet architecture is a lightweight neural network with fewer parameters that can more easily be transferred over a computer network (Iandola et al., 2016). Thus, in the future works, the proposed artificial intelligence may be able to operate even in microcontrollers. Moreover, since it will be embedded in a compact device, it may be able to diagnose simultaneously and provide information to the doctor. This is important for data privacy protection.

During transfer learning studies, firstly, the convolution layers of ResNet50, DenseNet161, and SequeezeNet1_1 architectures were frozen, and only fully connected layers were retrained. The learning rate was explored from 1.0×10^{-6} to 1.0 for all architectures, then it was stopped when loss diverged. Following the completion of the learning rate exploration process, the learning rate values for each architecture were similarly determined as 1.0×10^{-2}. The maximum number of epochs was 24.

In the fine-tuning process done for optimization, the frozen convolutional layers were retrained by unfreezing the layers. In the back-propagation process, the "stochastic gradient descent with restarts" (Loshchilov and Hutter, 2016) technique is used to find the minimum points on the loss function. As the iterations progress, the learning rate value is updated by repeating with the "cosine annealing" method and by expanding the period. Thanks to the dynamically updated learning rate, the minimum points on the "loss function" can be determined more accurately. To apply the "stochastic gradient descent with restarts" technique to the network trained at the transfer learning stage, the limits of the learning rate values should be determined. To determine the limit of the learning rate values, the learning rate exploration process of unfrozen networks was repeated and the appropriate learning rate interval to be studied was determined as 1.0×10^{-5} to 1.0×10^{-4}.

2.4. Performance evaluation criteria

The accuracy, sensitivity (recall), and precision (positive predictive value) parameters are the main parameters to determine the performance of the model. In the confusion matrix, there are four possible outcomes including true positive (TP), true negative (TN), false positive

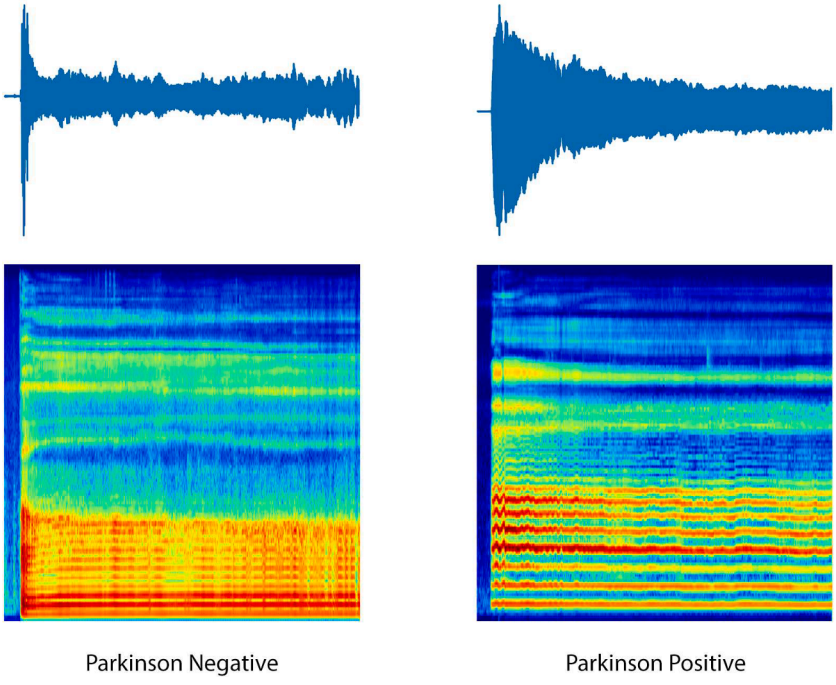

Fig. 5. The time-domain waveforms and Mel-spectrograms of the control group sample and PD positive sample.

(FP), false negative (FN) [68]. To evaluate the performance of the tailored convolutional neural networks the accuracy, recall, and precision coefficients were determined as followings.

$$Accuracy\% = \left(\frac{TP + TN}{TP + TN + FP + FN}\right) \times 100 \quad (1)$$

$$Sensivity\% = \left(\frac{TP}{TP + FN}\right) \times 100 \quad (2)$$

$$Precision\% = \left(\frac{TP}{TP + FP}\right) \times 100 \quad (3)$$

3. Results and discussion

The training process consists of two basic stages as followings; i) different architectures in various depths including ResNet, DenseNet and SequeezeNet were examined and all convolutional layers were frozen and retrained ii) the most appropriate architecture was determined and fine-tuning process was performed for each architecture.

In the first stage of the transfer learning stage, fully connected layers of ResNet, DenseNet, and SqueezeNet architectures, previously trained by ImageNet datasets with different layer numbers, were retrained and structured to distinguish between PD-positive and PD-negative cases. Then, fine-tuning on ResNet-50, DenseNet-161, and SqueezeNet1_1 architectures by considering both computational efficiency and the performance of networks in similar medical studies were conducted. As a result of 24 Epoch for DenseNet161 and ResNet50, and of 18 Epoch for SqueezeNet1_1, the accuracy, and validation loss metrics were calculated after each epoch and the results were plotted as depicted in Fig. 6. While validation loss values for DensNet161 and ResNet50 architectures appeared to decrease and had a stable structure by reaching saturation, this value for SqueezeNet1_1 showed a relatively unstable behavior. The accuracy values of DensNet161 and ResNet50 architectures appeared to be close to each other whereas SqueezeNet1_1 was calculated much lower than the others.

The confusion matrices calculated for both validation dataset and testing datasets were depicted in Fig. 7. The results revealed that the DenseNet-161 architecture performs better than other architectures with the following graphs (Fig. 7). It was observed from the confusion matrix of DenseNet-161 architecture calculated for the testing dataset reported in Fig. 7d that the proposed model can detect 176 out of 200 patients with PD as PD-positive, 183 out of 200 PD-negative cases as PD-negative. On the other hand, it was seen that it misclassified 17 PD-negative cases as PD-positive case.

In Table 2, a comparison of performance metrics of developed CNN model for different architectures both for validation data and independent testing data. DenseNet-161 architecture shows the best performance in accuracy, recall, and precision metrics. Additionally, it should be noted that all trained networks presented high recall (sensitivity) values. The results reveals that DenseNet-161 presents the highest accuracy value as 89.75% amongst the other architectures. Thanks to the relatively faster nature of the CNN than other techniques, the CNN model based on DenseNet-161 obtained a test time of 0.029±0.028 s whereas a train time of 0.003±0.005, which are satisfactory and acceptable periods for a medical examination.

When similar studies in the literature have been examined, that have used mPower data set in the training of artificial intelligence model developed to be used in PD diagnosis from voice-samples, it is seen that some of them have reached much higher accuracy values (Schwab & Karlen, 2019; Singh & Xu, 2020; Wroge et al., 2018; Zhang et al., 2018). Most of the these models consider the same voice samples both in the training and the testing data set. As a result, this bias leads to over-optimistic performance results since the model has learned to detect the characteristics of certain speakers and predicted the label in the test set.

4. Conclusions

An applicable PD identification application with high sensitivity and accuracy has been still investigating. The main trouble faced in the diagnosis of Parkinson's disease is that there is no single non-winvasive clinical screening test to detect PD at an early-stage. Experts can use the voice abnormalities caused by the weakening of the muscles controlling the vocal cords in advanced-stage Parkinson's patients as a manifest of PD disease. However, in some PD-cases, it is difficult to make a physical diagnosis by experts from the voice of patients, so they are more likely to be unable to diagnose or misdiagnose. Additionally, there is great need an innovative technology that assists the physician to cure and avoid spreading the disease to other cells of the brain. Bearing this in mind, in this study an easy-to-apply, sensitive, and effective transfer learning-based deep CNN model was proposed for diagnosis of PD. It was investigated how voice dataset from one large dataset using fine-tuning approaches of transfer learning model could enhance the identification of PD. Also, three different fine-tuning architectures were analyzed, evaluated, and compared to determine the most suitable one for PD diagnosis. The results revealed that the proposed deep CNN model based on transfer learning with a fine-tuning approach provides an acceptable detection of PD with an accuracy of 91.17%. One of the most important aspects of the study that will contribute to the literature is that it can detect PD disease in a wide spectrum by examining only the voice features of different levels of PD patients. Furthermore, in future studies, by implementing the proposed transfer learning model to a smart electronic device for personal usage, it will be possible to detect PD accurately and rapidly without interrupting the user. In the future prospect, the authors have been planning to develop a model that can make a multi-

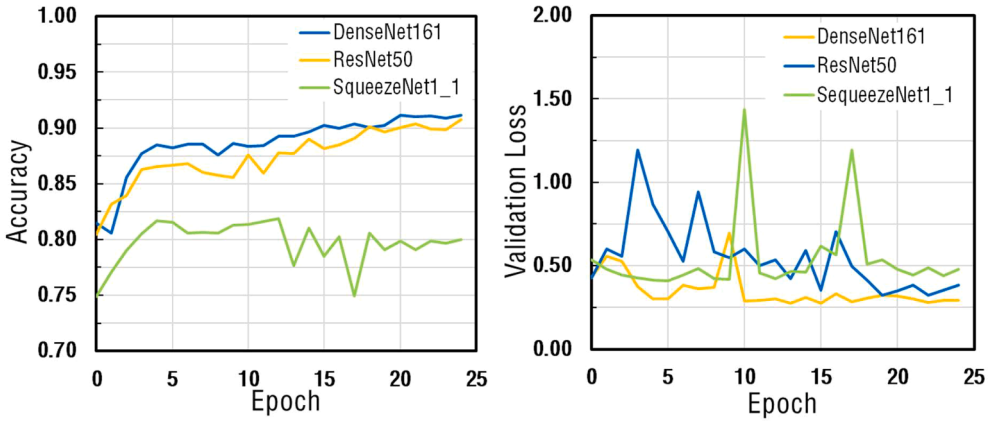

Fig. 6. The comparison of performance graphs of different metrics for DenseNet-161, ResNet-50, and SqueezeNet1_1 architectures.

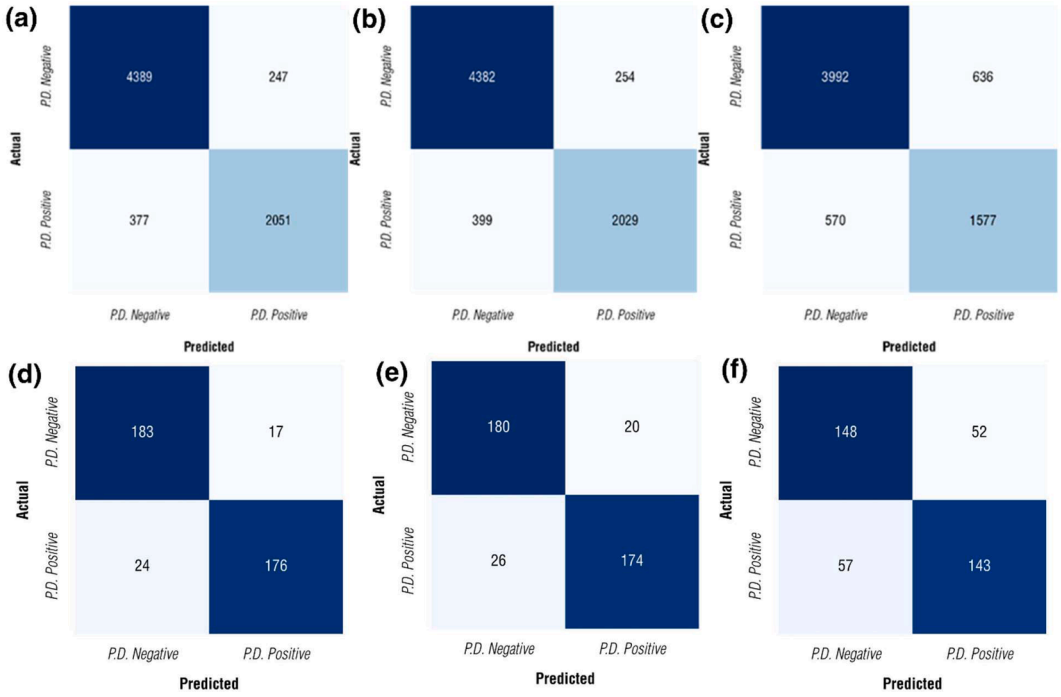

Fig. 7. Confusion matrices of optimized model for (a, d) DenseNet-161 (b, e) ResNet-50 (c, f) SqueezeNet1_1 architectures (a, b,c represents the calculated confusion matrices for the validation dataset, and d,e,f represents the calculated confusion matrices for the testing dataset).

Table 2
Comparison of calculated performance metrics for different architectures.

PerformanceMetrics	Architecture DenseNet-161			ResNet-50			SqueezeNet1_1		
	Accuracy	*Sensitivity*	*Precision*	*Accuracy*	*Sensitivity*	*Precision*	*Accuracy*	*Sensitivity*	*Precision*
Validation	91.17	84.47	89.25	90.76	83.57	88.87	81.85	73.43	73.68
Testing	**89.75**	**91.50**	**88.40**	88.50	90.00	87.30	72.75	74.00	72.19

classification of similar diseases and can score the PD-stage. The clinical impact is the possibility of the physician to utilize the developed model during the in-clinic assessment, while the machine-learning area impact is related to the proposed transfer learning model. Furthermore, the success of the proposed model would also imply an enhancement in the life quality of patients and a cost reduction for the national health system. Undoubtedly, there is a technological and scientific challenge to develop and disseminate expert systems for these tasks. However, they will be able to be incorporated into protocols by neurological units.

5. Ethical standards

This article does not contain any studies with human participants or animals performed by any of the authors.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgment

This publication was supported by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University, Alkharj, Saudi Arabia.

References

Arena, J. E., & Stoessl, A. J. (2016). Optimizing diagnosis in Parkinson's disease: Radionuclide imaging. *Parkinsonism & Related Disorders, 22*, 47–51. https://doi.org/10.1016/j.parkreldis.2015.09.029

Afonso, L. C. S., Rosa, G. H., Pereira, C. R., Weber, S. A. T., Hook, C., Albuquerque, V. H. C., et al. (2019). A recurrence plot-based approach for Parkinson's disease identification. *Future Generation Computer Systems, 94*, 282–292. https://doi.org/10.1016/j.future.2018.11.054

ALAC. 2020. Apple Lossless Audio Codec. Retrieved from https://macosforge.github.io/alac/. Accessed December 15, 2020.

Almeida, J. S., Rebouças Filho, P. P., Carneiro, T., Wei, W., Damaševičius, R., Maskeliūnas, R., et al. (2019). Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. *Pattern Recognition Letters, 125*, 55–62. https://doi.org/10.1016/j.patrec.2019.04.005

Berus, L., Klancnik, S., Brezocnik, M., & Ficko, M. (2019). Classifying Parkinson's disease based on acoustic measures using artificial neural networks. *Sensors, 19*(1), 16. https://doi.org/10.3390/s19010016

Bernardo, L. S., Quezada, A., Munoz, R., Maia, F. M., Pereira, C. R., Wu, W., et al. (2019). Handwritten pattern recognition for early Parkinson's disease diagnosis. *Pattern Recognition Letters, 125*, 78–84. https://doi.org/10.1016/j.patrec.2019.04.003

Blumin, J. H., Pcolinsky, D. E., & Atkins, J. P. (2004). Laryngeal findings in advanced Parkinson's disease. *Annals of Otology, Rhinology & Laryngology: SAGE Journals, 113*(4), 253–258.

Bugalho, P., & Viana-Baptista, M. (2013). REM sleep behavior disorder and motor dysfunction in Parkinson's disease–a longitudinal study. *Parkinsonism & Related Disorders, 19*(12), 1084–1087. https://doi.org/10.1016/j.parkreldis.2013.07.017

Caekebeke, J. F., Jennekens-Schinkel, A., Van der Linden, M. E., Buruma, O. J., & Roos, R. A. (1991). The interpretation of dysprosody in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 54*(2), 145–148. https://doi.org/10.1136/jnnp.54.2.145

Caliskan, A., Badem, H., Basturk, A., & Yuksel, M. E. (2017). Diagnosis of the parkinson disease by using deep neural network classifier. *Journal of Electrical and Electronic Engineering, 17*(2), 3311–3318.

Chenausky, K., MacAuslan, J., & Goldhor, R. (2011). Acoustic analysis of PD speech. *Parkinson's Disease, 2011*, 1–13. https://doi.org/10.4061/2011/435232

Choi, H., Ha, S., Im, H. J., Paek, S. H., & Lee, D. S. (2017). Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage: Clinical, 16*, 586–594. https://doi.org/10.1016/j.nicl.2017.09.010

Coates, A., Ng, A., & Lee, H. (2011). An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 215–223).

Correia, J., Trancoso, I., Raj, B. (2019). In-the-Wild End-to-End Detection of Speech Affecting Diseases. In 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 734–741.IEEE.

de Souza, J. W. M., Alves, S. S. A., Rebouças, E.d. S., Almeida, J. S., & Rebouças Filho, P. P. (2018). A new approach to diagnose parkinson's disease using a structural cooccurrence matrix for a similarity analysis. *Computational Intelligence and Neuroscience, 2018*, 1–8. https://doi.org/10.1155/2018/7613282

de Souza, R. W. R., Silva, D. S., Passos, L. A., Roder, M., Santana, M. C., Pinheiro, P. R., et al. (2021). Computer-assisted Parkinson's disease diagnosis using fuzzy optimum-path forest and Restricted Boltzmann Machines. *Computers in Biology and Medicine, 131*, 104260. https://doi.org/10.1016/j.compbiomed.2021.104260

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., et al. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-Rhino-Laryngology, 258*(2), 77–82.

Delrobaei, M., Memar, S., Pieterman, M., Stratton, T. W., McIsaac, K., & Jog, M. (2018). Towards remote monitoring of Parkinson's disease tremor using wearable motion capture systems. *Journal of the Neurological Sciences, 384*, 38–45. https://doi.org/10.1016/j.jns.2017.11.004

Erdogdu Sakar, B., Serbes, G., & Sakar, C. O. (2017). Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. *PloS One, 12*, Article e0182428.

Fastai. (2020). Retrieved from https://www.fast.ai/. Accessed December 15, 2020.

Grósz, T., Busa-Fekete, R., Gosztolya, G., & Tóth, L. (2015). Assessing the degree of nativeness and Parkinson's condition using Gaussian processes and deep rectifier neural networks. In *Proceedings of Interspeech* (pp. 1339–1343).

Gupta, D., Sundaram, S., Khanna, A., Hassanien, A. E., & De Albuquerque, V. H. C. (2018). Improved diagnosis of Parkinson's disease using optimized crow search algorithm. *Computers & Electrical Engineering, 68*, 412–424. https://doi.org/10.1016/j.compeleceng.2018.04.014

Gupta, D., Julka, A., Jain, S., Aggarwal, T., Khanna, A., Arunkumar, N., et al. (2018). Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. *Cognitive Systems Research, 52*, 36–48. https://doi.org/10.1016/j.cogsys.2018.06.006

Hanning. (2020) Hanning Function. Retrieved from https://mathworld.wolfram.com/HanningFunction.html. Accessed December 15, 2020.

Haq, Amin Ul, Li, Jian Ping, Memon, Muhammad Hammad, khan, Jalaluddin, Malik, Asad, Ahmad, Tanvir, et al. (2019). Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. *IEEE Access, 7*, 37718–37734.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition ,770–778.

Hin, S., Shukla, G., Goyal, V., Singh, S., Aggarwal, V., & Behari, M. (2008). Clinical speech impairment in Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy. *Neurol India, 56*(2), 122. https://doi.org/10.4103/0028-3886.41987

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).

Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv: preprint arXiv:1602.07360.

ImageNET. (2020). Retrieved from http://www.image-net.org/papers/imagenet_cvpr09.bi. Accessed December 15, 2020.

Júnior, Eugênio Peixoto, Delmiro, Italo L. D., Magaia, Naercio, Maia, Fernanda M., Hassan, Mohammad Mehedi, Albuquerque, Victor Hugo C., et al. (2020). Intelligent sensory pen for aiding in the diagnosis of parkinson's disease from dynamic handwriting analysis. *Sensors, 20*(20), 5840. https://doi.org/10.3390/s20205840

Kostikis, N., Hristu-Varsakelis, D., Arnaoutoglou, M., & Kotsavasiloglou, C. (2015). A smartphone-based tool for assessing parkinsonian hand tremor. *IEEE Journal of Biomedical and Health Informatics, 196*(6), 1835–1842. https://doi.org/10.1109/JBHI.2015.2471093

Lahmiri, S., & Shmuel, A. (2019). Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. *Biomedical Signal Processing and Control, 49*, 427–433. https://doi.org/10.1016/j.bspc.2018.08.029

Lauraitis, Andrius, Maskeliūnas, Rytis, & Damaševičius, Robertas (2018). ANN and fuzzy logic based model to evaluate huntington disease symptoms. *Journal of Healthcare Engineering, 2018*, 1–10. https://doi.org/10.1155/2018/4581272

Librosa. (2020). Retrieved from https://librosa.org/doc/latest/index.html. Accessed December 15, 2020.

Loshchilov, I., Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. arXiv: preprint arXiv:1608.03983.

Mallela, J., Illa, A., Suhas, B.N., Udupa, S., Belur, Y., et al. (2020). Voice based classification of patients with Amyotrophic Lateral Sclerosis, Parkinson's Disease and Healthy Controls with CNN-LSTM using transfer learning. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6784–6788.IEEE.

Martens, H., Van Nuffelen, G., Wouters, K., & De Bodt, M. (2016). Reception of communicative functions of prosody in hypokinetic dysarthria due to Parkinson's disease. *Journal of Prosthetic Dentistry, 6*(1), 219–229. https://doi.org/10.3233/JPD-150678

Millian-Morell, Lymarie, Lopez-Alburquerque, Tomas, Rodriguez-Rodriguez, Andrea, Gomez-Nieto, Ricardo, Carro, Juan, Meilan, Juan J. G., et al. (2018). Relations between sensorimotor integration and speech disorders in Parkinson's disease. *Current Alzheimer Research, 15*(2), 149–156. https://doi.org/10.2174/1567205014666170829103019

Moetesum, M., Siddiqi, I., Vincent, N., & Cloppet, F. (2019). Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease. *Pattern Recognition Letters, 121*, 19–27. https://doi.org/10.1016/j.patrec.2018.04.008

mpower App store. (2020) Retrieved from https://itunes.apple.com/us/app/parkinson-mpower-study-app/id972191200?mt=8. Accessed September 10, 2020.

Naseer, A., Rani, M., Naz, S., Razzak, M. I., Imran, M., & Xu, G. (2020). Refining Parkinson's neurological disorder identification through deep transfer learning. *Neural Computing and Applications, 32*(3), 839–854. https://doi.org/10.1007/s00521-019-04069-0

New, Anneliese B., Robin, Donald A., Parkinson, Amy L., Eickhoff, Claudia R., Reetz, Kathrin, Hoffstaedter, Felix, et al. (2015). The intrinsic resting state voice network in Parkinson's disease. *Human Brain Mapping, 36*(5), 1951–1962.

Pahuja, G., Nagabhushan, T.N. (2018). A Comparative Study of Existing Machine Learning Approaches for Parkinson's Disease Detection. IETE Journal of Research, 1-11. https://doi.org/10.1080/03772063.2018.1531730.

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage, 61*(1), 314–322. https://doi.org/10.1016/j.neuroimage.2012.02.068

Parkinson, James (2002). An essay on the shaking palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences, 14*(2), 223–236. https://doi.org/10.1176/jnp.14.2.223

Pawlukowska, Wioletta, Gołąb-Janowska, Monika, Safranow, Krzysztof, Rotter, Iwona, Amernik, Katarzyna, Honczarenko, Krystyna, et al. (2015). Articulation disorders and duration, severity and L-dopa dosage in idiopathic Parkinson's disease. *Neurologia i Neurochirurgia Polska, 49*(5), 302–306. https://doi.org/10.1016/j.pjnns.2015.07.002

Peker, Musa, Şen, Baha, & Delen, Dursun (2015). Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm. *Journal of Healthcare Engineering, 6*(3), 281–302.

Peker, M. (2016). A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM. *Journal of Medical Systems, 40*(5), 116. https://doi.org/10.1007/s10916-016-0477-6

Pereira, Clayton R., Pereira, Danilo R., Silva, Francisco A., Masieiro, João P., Weber, Silke A. T., Hook, Christian, et al. (2016). A new computer vision-based approach to aid the diagnosis of Parkinson's disease. *Computer Methods and Programs in Biomedicine, 136*, 79–88. https://doi.org/10.1016/j.cmpb.2016.08.005

Pereira, Clayton R., Pereira, Danilo R., Rosa, Gustavo H., Albuquerque, Victor H. C., Weber, Silke A. T., Hook, Christian, et al. (2018). Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification. *Artificial Intelligence in Medicine, 87*, 67–77. https://doi.org/10.1016/j.artmed.2018.04.001

Pereira, C. R., Pereira, D. R., Weber, S. A., Hook, C., de Albuquerque, V. H. C., & Papa, J. P. (2019). A survey on computer-assisted Parkinson's disease diagnosis. *Artificial Intelligence in Medicine, 95*, 48–63. https://doi.org/10.1016/j.artmed.2018.08.007

Polat, Ç., Karaman, O., Karaman, C., Korkmaz, G., Balcí, M. C., & Kelek, S. E. COVID-19 diagnosis from chest X-ray images using transfer learning: Enhanced performance by debiasing dataloader. Journal of X-Ray Science and Technology, 29(1), 19–36. https://doi.org//10.3233/XST-200757.

Ramig, Lorraine, Halpern, Angela, Spielman, Jennifer, Fox, Cynthia, & Freeman, Katherine (2018). Speech treatment in Parkinson's disease: Randomized controlled trial (RCT). *Movement Disorders, 33*(11), 1777–1791.

Reeve, A., Simcox, E., & Turnbull, D. (2014). Ageing and Parkinson's disease: Why is advancing age the biggest risk factor? *Ageing Research Reviews, 14*, 19–30. https://doi.org/10.1016/j.arr.2014.01.004

[dataset] Sage Bionetworks. (2020) Retrieved from https://www.synapse.org/mPower. Accessed September 15, 2020.

Sakar, C. O., & Kursun, O. (2010). Telediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of Medical Systems, 34*(4), 591–599. https://doi.org/10.1007/s10916-009-9272-y

Sapir, S. (2014). Multiple factors are involved in the dysarthria associated with Parkinson's disease: A review with implications for clinical practice and research. *Journal of Speech, Language, and Hearing Research, 57*(4), 1330–1343. https://doi.org/10.1044/2014_JSLHR-S-13-0039

Schwab, P., & Karlen, W. (2019). PhoneMD: Learning to diagnose Parkinson's disease from smartphone data. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*, 1118–1125.

Shirvan, R. A., & Taham, E. (2011). Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method. In *2011 18th Iranian Conference of Biomedical Engineering (ICBME)* (pp. 278–283). https://doi.org/10.1109/ICBME.2011.6168572

Singh, Sanjana, & Xu, Wenyao (2020). Robust detection of parkinson's disease using harvested smartphone voice data: A telemedicine approach. *Telemedicine and e-Health, 26*(3), 327–334.

Singh, G., & Samavedham, L. (2015). Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: A case study on early-stage

diagnosis of Parkinson disease. *Journal of Neuroscience Methods, 256*, 30–40. https://doi.org/10.1016/j.jneumeth.2015.08.011

Skodda, Sabine (2011). Aspects of speech rate and regularity in Parkinson's disease. *Journal of the Neurological Sciences, 310*(1-2), 231–236. https://doi.org/10.1016/j.jns.2011.07.020

Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions on Biomedical Engineering, 59*(5), 1264–1271. https://doi.org/10.1109/TBME.2012.2183367

Upadhya, S. S., Cheeran, A. N., & Nirmal, J. H. (2018). Thomson Multitaper MFCC and PLP voice features for early detection of Parkinson disease. *Biomedical Signal Processing and Control, 46*, 293–301. https://doi.org/10.1016/j.bspc.2018.07.019

Van Stan, J. H., Park, S. W., Jarvis, M., Mehta, D. D., Hillman, R. E., & Sternad, D. (2017). Measuring vocal motor skill with a virtual voice-controlled slingshot. *The Journal of the Acoustical Society of America, 142*(3), 1199–1212. https://doi.org/10.1121/1.5000233

Vasquez-Correa, Juan Camilo, Arias-Vergara, Tomas, Orozco-Arroyave, J. R., Eskofier, Bjorn, Klucken, Jochen, & Noth, Elmar (2019). Multimodal assessment of Parkinson's disease: A deep learning approach. *IEEE Journal of Biomedical and Health Informatics, 23*(4), 1618–1630.

Wroge, T.J., Özkanca, Y., Demiroglu, C., Si, D., Atkins, D.C., Ghomi, R.H. (2018). Parkinson's disease diagnosis using machine learning and voice. In 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2018,1–7.

Xia, Y., Yao, Z., Lu, Y., Zhang, D., & Cheng, N. (2018). A machine learning approach to detecting of freezing of gait in parkinson's disease patients. *Journal of Medical Imaging and Health Informatics, 8*(4), 647–654. https://doi.org/10.1166/jmihi.2018.2379

Xu, C., He, J., Zhang, X., Wang, C., & Duan, S. (2018). Template-matching-based detection of freezing of gait using wearable sensors. *Procedia Computer Science, 129*, 21–27. https://doi.org/10.1016/j.procs.2018.03.038

Zhang, H., Song, C., Wang, A., Xu, C., Li, D., & Xu, W. (2015). Pdvocal: Towards privacy-preserving parkinson's disease detection using non-speech body sounds. In *The 25th Annual International Conference on Mobile Computing and Networking* (pp. 1–16). https://doi.org/10.1145/3300061.3300125

Zhang, H., Wang, A., Li, D., Xu, W. (2018). Deepvoice: A voiceprint-based mobile health framework for parkinson's disease identification. In 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 214-217. IEEE.

Zhang, Y. N. (2017). Can a smartphone diagnose parkinson disease? a deep neural network method and telediagnosis system implementation. *Parkinson's Disease, 2017*, 1–11. https://doi.org/10.1155/2017/6209703

Zhang, Z. (2016). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America, 140*(4), 2614–2635. https://doi.org/10.1121/1.4964509